

STUDI IMPLEMENTASI SISTEM *BIG DATA* UNTUK MENDUKUNG KEBIJAKAN KOMUNIKASI DAN INFORMATIKA

Study on Implementation of Big Data System for Supporting Communication and Informatics Policy

Dita Kusumasari¹ dan Onny Rafizan²

^{1,2)} Puslitbang Aptika dan IKP, Kementerian Komunikasi dan Informatika
Jl. Medan Merdeka Barat No.9, Jakarta
E-mail: dita001@kominfo.go.id¹, onny002@kominfo.go.id²

Naskah diterima tanggal 16 September 2017, direvisi tanggal 27 November 2017, disetujui tanggal 15 Desember 2017

Abstract

Media Monitoring became one of the tasks and functions of the Ministry of Communications and Information Technology, in accordance with Presidential Instruction No. 9/2015. The process should be done in short time while maintaining or even improving the accuracy of analysis of the media. Therefore Big Data technology becoming a promising solution, related abilities of Big Data to process variety of data in a large scale, also provide accurate reports for stakeholders. This study adopted the Modified Waterfall method that commonly used in manufacture of software. This method is expected to explores and creates an appropriate recommendation. Big Data Implementation process will require time, cost, and human resources, so that stakeholders and related user is expected to prepare the implementation process properly in order to be effective.

Keywords : *Media Monitoring, Big Data, Modified Waterfall method, Implementation*

Abstrak

Monitoring media menjadi salah satu tugas dan fungsi Kementerian Komunikasi dan Informatika, sesuai dengan Instruksi Presiden No. 9 Tahun 2015. Proses dalam monitoring media tersebut harus dapat diselesaikan dalam waktu cepat tanpa mengurangi akurasi dari analisis terhadap media tersebut. Oleh karena itu teknologi *Big Data* menjadi salah satu solusi yang menjanjikan, terkait sifatnya yang mampu mengolah data dalam skala yang sangat besar dan variatif serta menyajikan laporan yang akurat untuk digunakan oleh pemangku kebijakan. Penelitian ini mengadopsi metode *Modified Waterfall* yang biasa digunakan dalam pembuatan *software*. Metode ini diharapkan dapat mendalami, mengeksplorasi dan menghasilkan alternatif rekomendasi yang sesuai. Proses pengimplementasian *Big Data* akan membutuhkan biaya, SDM, dan waktu yang tidak sedikit, sehingga pemangku kebijakan dan *user* terkait diharapkan dapat lebih mempersiapkan dengan baik agar proses pengimplementasian dapat berjalan efektif dan maksimal.

Kata Kunci : *Monitoring Media, Big Data, Metode Modified Waterfall, Implementasi*

PENDAHULUAN

Perkembangan dalam penggunaan teknologi informasi dan komunikasi telah merubah cara berkomunikasi, khususnya dalam penyebaran informasi. Mulanya metode komunikasi dan distribusi informasi hanya terbatas pada media tertulis (kertas, surat) dan media elektronik (radio, televisi, dan telepon) sehingga informasi yang beredar masih sangat terbatas, baik dari skala informasi yang beredar

maupun area yang mampu dijangkau, terlebih pada distribusi informasi lintas Negara. Masyarakat di suatu Negara tidak dapat dengan mudah memperoleh dan mengakses informasi yang berkaitan dengan Negara lain, begitu pun sebaliknya. Berkembangnya internet di era teknologi yang maju memungkinkan peredaran informasi yang semakin banyak, cepat, dan hampir tidak terbatas oleh ruang dan waktu. Hal ini menyebabkan informasi yang beredar dari hari ke hari menjadi sangat banyak, mencakup berbagai informasi dalam bidang

sosial, politik, ekonomi, teknologi, ilmu pengetahuan, pangan, dan lain sebagainya.

Kondisi serupa juga terjadi dalam skala pemerintahan. Beberapa tahun terakhir telah muncul himbauan untuk mengalihkan informasi/dokumen fisik menjadi data elektronik. Pelaksanaan himbauan tersebut salah satunya bertujuan untuk memudahkan pengelolaan dan penggunaan data saat dibutuhkan. Di satu sisi hal ini dapat segera dilaksanakan jika jenis dan format dari seluruh data yang ada seragam. Namun di sisi lain, semakin beragam jenis dan format data, serta semakin besar jumlah dan variasi data akan memakan waktu yang lama untuk mengolah data – data tersebut menjadi informasi yang terstruktur rapi dan dapat dibaca.

Tidak dapat dipungkiri, informasi merupakan senjata yang sangat penting saat ini. Organisasi seperti pemerintah tentunya memiliki data yang sangat beragam dan dengan jumlah yang sangat banyak. Untuk dapat mengambil keputusan yang tepat, dalam hal ini pemerintah selaku pembuat kebijakan, memerlukan cara yang efektif untuk mengelola data – data tersebut menjadi informasi yang berguna sebagai pertimbangan dalam membuat keputusan.

Media menjadi salah satu bagian yang menjadi perhatian pemerintah, salah satunya tertuang dalam Instruksi Presiden No. 9 Tahun 2015 tentang Pengelolaan Komunikasi Publik terkait tugas dan fungsi Kementerian Komunikasi dan Informatika, yaitu:

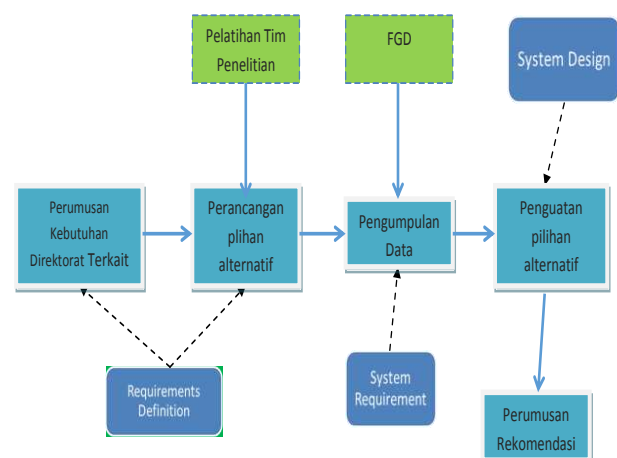
1. Melakukan kajian terhadap data dan informasi yang disampaikan kementerian dan lembaga pemerintah non kementerian;
2. Melakukan Media Monitoring dan menganalisis konten media terkait dengan kebijakan dan program pemerintah;
3. Menyusun narasi tunggal terkait dengan kebijakan dan program pemerintah kepada publik sesuai arahan Presiden.

Pengelolaan data yang beragam dengan jumlah yang sangat besar membutuhkan suatu

cara yang efektif untuk mengolahnya, terlebih jika informasi yang dihasilkan dari data tersebut dibutuhkan untuk membantu membuat keputusan bagi pemangku kebijakan. Diperlukan cara yang cepat dan tepat untuk dapat mengolah data tersebut menjadi informasi. Oleh karena itu prinsip *Big Data* sangat cocok diterapkan, dimana prinsip *Big Data* yaitu untuk dapat mengelola data yang sangat banyak dan beragam, serta mengolahnya menjadi informasi yang dibutuhkan dalam waktu yang sangat singkat. Dengan menemukan pilihan alternatif implementasi *Big Data* yang cocok diterapkan pada Kementerian Komunikasi dan Informatika, khususnya pada bagian Media Monitoring, diharapkan dapat membantu *user* terkait dalam membuat keputusan melalui implementasi *Big Data*.

Metode Penelitian

Penelitian ini menggunakan pendekatan secara kualitatif untuk mendalami kebutuhan yang ada. Dalam menemukan alternatif implementasinya, penelitian ini menggunakan metode *Modified Waterfall*. Karena batasan penelitian ini tidak mencakup pembuatan dan implementasi sistem, maka jika mengacu pada *Modified Waterfall*, tahap yang dilakukan hanya terbatas hingga tahap *System Design*. Oleh karena itu, dengan mengacu pada metode tersebut, serta mempertimbangkan batasan penelitian yang ditetapkan, maka tahapan penelitian disesuaikan seperti dalam Gambar 1.



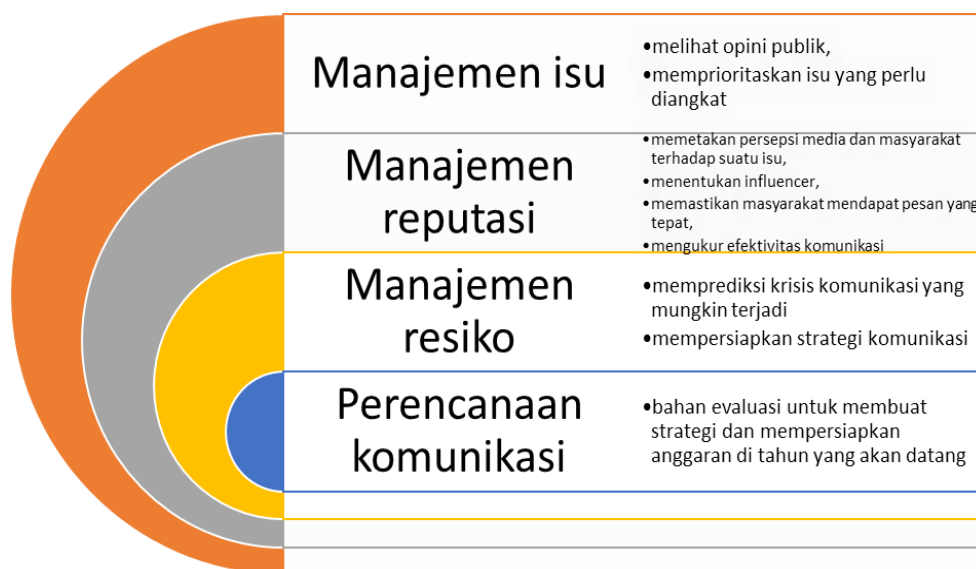
Gambar 1. Pendekatan Metode Studi

Studi Literatur

Pemerintahan yang baik hanya dapat terjadi bila pengambilan keputusan didasarkan pada informasi yang memadai serta penilaian yang independen (Sullivan dalam Subiakto, 2014: 248). Hal demikian dapat dicapai dengan adanya informasi yang faktual dan terpercaya, yang hanya dapat diperoleh dari pers yang bebas, yang berfungsi sebagai *watchdog* masyarakat atas pemerintah. Dalam Negara yang menjamin kebebasan pers dan informasi, pemerintah harus siap terhadap keadaan tersebut. Disinilah peran lembaga informasi pemerintah, yakni sebagai pusat komunikasi dengan publik dalam menjelaskan rencana dan program pemerintah terhadap masyarakat, sehingga masyarakat dapat memahami pengaruh dan peranan kebijakan tersebut bagi kehidupan mereka.

Pemerintah yang memiliki begitu banyak informasi untuk disampaikan memerlukan cara efektif dan peran juru bicara pemerintah untuk menyampaikan informasi-informasi tersebut kepada publik. Secara teoritis humas pemerintah atau *Government Public Relations (GPR)* mempunyai tugas untuk menjelaskan dampak program dan kebijakan pemerintah terhadap warganya, termasuk isu-isu kontroversial yang beredar.

Pemerintahan Kabinet Kerja sedang merancang agar data yang melimpah dalam media konvensional maupun media baru tersebut dapat dimanfaatkan sebagai bentuk opini publik yang dapat membentuk citra positif bagi pemerintah. Melalui Instruksi Presiden Nomor 9 Tahun 2015 tentang Pengelolaan Komunikasi Publik, Kementerian Komunikasi dan Informatika telah ditugaskan untuk melakukan monitoring media.



Gambar 2. Elemen Kehumasan

Monitoring Media

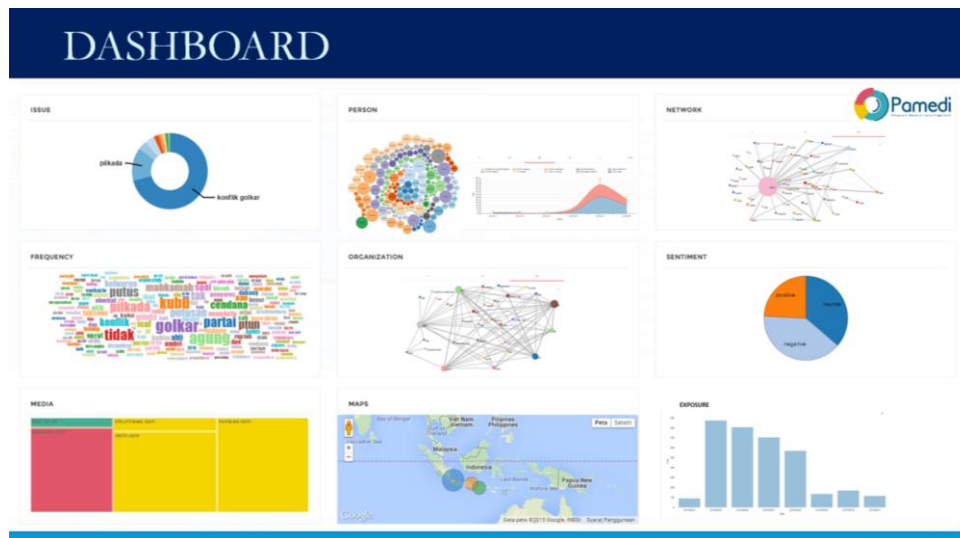
Menurut Direktorat Jenderal Informasi dan Komunikasi Publik (Ditjen IKP), yang dimaksud dengan Monitoring Media adalah kegiatan memantau konten-konten yang beredar di media, baik cetak, penyiaran maupun *online* dan media sosial. Adapun untuk memonitor media, dengan mengacu pada elemen kehumasan seperti ditunjukkan pada Gambar 2.

Data-data yang terdapat pada Direktorat Pengelolaan dan Penyediaan Informasi (PPI) bersifat tidak terstruktur dengan pertambahan jumlah data yang sangat cepat. Data tersebut harus dapat diolah dalam waktu yang singkat untuk menjadi informasi. Dibutuhkan metode statistik yang akurat dalam menampilkan informasinya, dimana hal tersebut dapat membantu pemangku kebijakan dalam

menerima informasi yang dibutuhkan untuk membuat keputusan.

Untuk itu, Direktorat PPI pada Ditjen IKP telah bekerjasama dengan PT. Delapan Sebelas Indonesia pada tahun 2015 hingga 2016, dan menghasilkan aplikasi pengelolaan Monitoring Media dengan *Big Data* yang

disebut dengan Pamedi (*Paques Media Intelligence*).



Gambar 3. Tampilan Dashboard Pamedi

Pamedi (*Paques Media Intelligence*)

Pamedi sebagai alat yang digunakan untuk monitoring media di Direktorat PPI, sudah mulai diterapkan pada tahun 2015 ini (lihat Gambar 3). Kemampuan Pamedi ini terbagi menjadi tiga¹:

- *Media Monitoring*: suatu monitoring topik atau isu yang ada pada media sosial maupun media berita secara sistematis, yang memungkinkan pengguna untuk mengakses secara *real-time* informasi dan alat pengukuran pada aplikasi sehingga memungkinkan pengguna untuk membuat keputusan yang lebih baik dan merespon suatu situasi yang penting.
- *Media Measurements & Analysis*: memungkinkan *user* untuk memaparkan grafik dengan analisis yang berkualitas berdasarkan hasil yang didapatkan pada aplikasi/ akun pengguna.

- *Expert Analysis & Reporting*: Pamedi juga memungkinkan untuk membuat suatu laporan yang dapat diubah sesuai kebutuhan lengkap dengan grafiknya.

Karakteristik *Big Data*

Suatu data dikategorikan sebagai "*Big Data*" bukan hanya karena jumlah data yang besar. Ada beberapa karakteristik yang membedakan *Big Data* dengan sistem lainnya.

Sistem *Big Data* memiliki **Volume** data yang sangat besar, dimana biasanya melebihi server biasa pada umumnya dan data ini akan bertambah terus tiap harinya. Besarnya data bisa mencapai lebih dari 100 TB dan biasanya disimpan di infrastruktur external (tidak di *maintain* sendiri).²

Selain itu *Big Data* juga memiliki data yang bervariasi (**Variety**), dengan format maupun jenis data yang sangat beragam

¹ Diambil dari Paparan PT. Delapan Sebelas Indonesia tentang Pamedi tanggal 2 September 2015

² Diambil dari Paparan Prof. Dr. Mochamad Ashari tentang *Big Data Industry and Academic Point of View*, dalam Konferensi Big Data Indonesia 2015.

sehingga memerlukan suatu proses khusus untuk dapat mengolahnya. *Big Data* juga harus dapat mengolah data tersebut dalam waktu yang sangat cepat (*Velocity*) agar data bisa bermanfaat tidak hanya karena informasi yang dihasilkan saja tetapi juga karena kecepatan yang dibutuhkan untuk mengolahnya menjadi informasi tersebut.

Karakteristik *Big Data* yang keempat adalah kebenaran akan data itu sendiri (*Veracity*). Informasi yang diolah dari data tersebut agar bisa menjadi informasi yang bermanfaat dan bisa dipercaya, kita juga harus melihat sumber data yang digunakan. Karenanya pada *Big Data*, kebenaran akan data menjadi salah satu hal yang harus diperhatikan.

HASIL DAN PEMBAHASAN

Sesuai Instruksi Presiden No. 9 Tahun 2015, tugas Pengelolaan Komunikasi Publik meliputi; Melakukan kajian terhadap data dan informasi yang disampaikan kementerian dan lembaga pemerintah non kementerian; serta Melakukan Media Monitoring dan menganalisis konten media terkait dengan kebijakan dan program pemerintah. Pada pelaksanaan kedua tugas tersebut, ditemukan “gap” antara kebijakan dan program pemerintah yang berjalan dengan isu pemberitaan media, sehingga perlu adanya upaya antisipasi oleh pemerintah. Oleh karena itu, monitoring media menjadi salah satu bagian penting dari Kementerian Komunikasi dan Informatika, tepatnya bagi Direktorat PPI, Ditjen IKP. Kebutuhan ini telah dituangkan dalam sebuah Proses Bisnis seperti ditunjukkan dalam Gambar 4.



Gambar 4. Proses Bisnis Monitoring Media

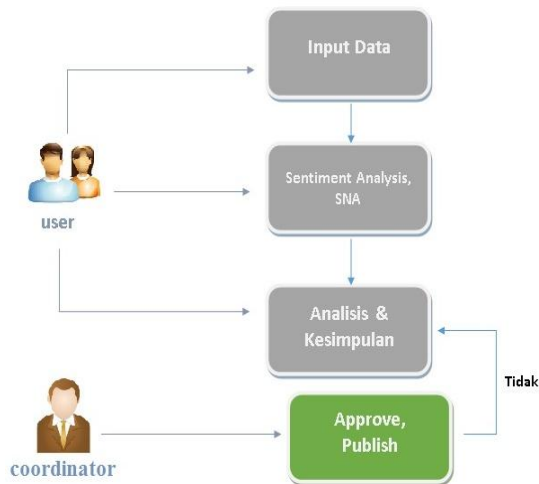
Proses monitoring media tersebut melibatkan 22 media dengan rincian: 12 Media cetak, 5 media *online*, dan 5 Media TV. Media yang dipilih sebagai sampel merupakan media yang representatif. Hal ini dikarenakan sifat media yang saat ini cenderung memiliki arah yang dapat dikelompokkan berdasarkan pemberitaannya. Secara garis besar proses monitoring media menghasilkan dua laporan

hasil analisis, yaitu Monitoring Isu Publik (MIP) dan *Media Content Analysis (MCA)*.

1. Monitoring Isu Publik (MIP)

MIP merupakan bagian dari proses monitoring media yang dilakukan melalui penelusuran *headline* berita. Kriteria *headline* yang dipantau yaitu *headline* berita yang mengandung informasi ataupun isu – isu yang

sedang marak diberitakan oleh media. Selama ini proses penentuan isu dilakukan dengan melihat subjektivitas pembaca berita. Jika menerapkan *Big Data* dengan metode *Sentiment Analysis* dapat mempercepat proses analisis berita dalam pengerjaan laporan MIP, dengan menerjemahkan sudut pandang seseorang kedalam bahasa mesin.



Gambar 5. Proses Bisnis MIP

Selain *Sentiment Analysis*, penggunaan metode *Social Network Analysis* (SNA) secara bersamaan juga mampu memperkaya hasil analisis, salah satunya dengan menyediakan informasi tentang keterkaitan antara satu isu di

suatu media dengan media lainnya. Proses bisnis baru dalam pengerjaan MIP dengan intervensi *Sentiment Analysis* dan SNA, ditunjukkan dalam Gambar 5.

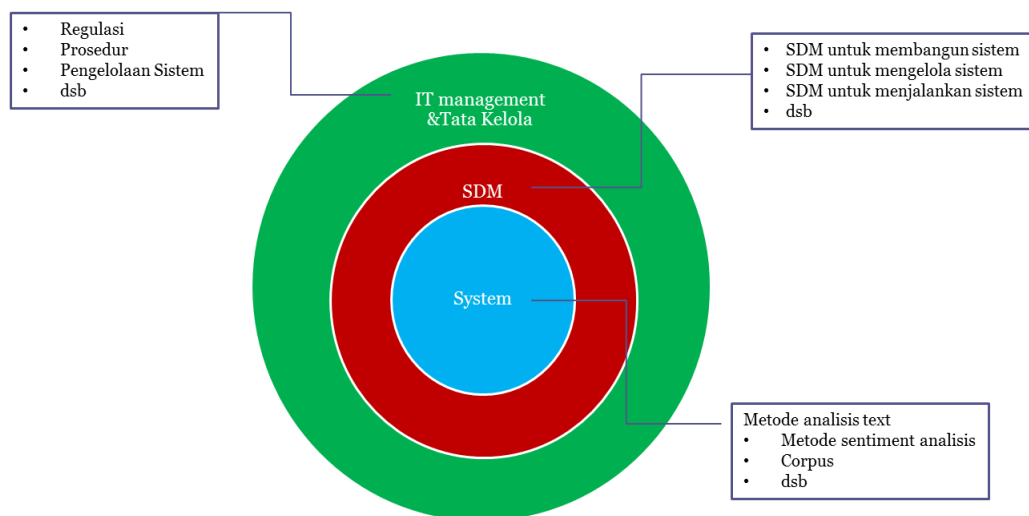
2. Media Content Analysis (MCA)

Media Content Analysis (MCA) merupakan proses monitoring media berdasarkan isi atau konten dari berita. Konten berita yang dilihat merupakan konten dari berita-berita dengan *headline* terbanyak. Sama halnya dengan MIP, penerapan metode *Sentiment Analysis* dan *Social Network Analysis* (SNA) dapat mempercepat proses analisis serta memperkaya hasil analisis berita.

Perbedaan dengan MIP terletak pada objek analisis. Jika analisis MIP hanya mengacu pada *headline* berita, maka MCA dihasilkan dari analisis isi suatu berita pada media.

Tahapan Implementasi *Big Data*

Karena *Big Data* mencakup area yang sangat luas, maka implementasinya dalam penelitian ini akan dikategorikan menjadi tiga tahap, yaitu *IT Management & Tata Kelola*, SDM, dan Sistem, seperti ditunjukkan dalam Gambar 6.



Gambar 6. Tahapan Implementasi *Big Data*

***IT Management* dan Tata Kelola**

a. *IT Management*

Gartner menggambarkan dimensi *Big Data* sebagai 3V yaitu *Volume*, *Velocity*, *Variety* (Gartner, 2009). Seiring dengan perkembangannya, *Big Data* tidak hanya mencakup 3V namun berkembang ke 5V yaitu *Volume*, *Velocity*, *Value*, *Veracity*, dan *Variety*. Secara teknis, *Big Data* merupakan sekumpulan data yang besar baik terstruktur, semi, maupun tidak terstruktur sehingga tidak bisa diolah menggunakan perangkat *database relational* biasa (Nitin Sawant, 2013).

Data-data yang muncul mempunyai peluang yang mampu memberikan sebuah petunjuk kebijakan tanpa pernah disadari sebelumnya (Milton, 2009). *Big Data* merupakan tren teknologi untuk melakukan pendekatan baru dalam memahami dunia dan membuat keputusan bisnis (John F. O., 2013). Keputusan-keputusan ini dibuat berdasarkan data dalam volume yang sangat besar terstruktur, tidak terstruktur dan kompleks (misalnya *tweet*, video, transaksi komersial).

Menurut Bill Schmarzo, proses integrasi *Big Data* di dalam sebuah enterprise mempunyai mempunyai indek kematangan bisnis yang terdiri dari beberapa fase sebagai berikut (Bill, 2013):

1. *Business Monitoring*.
2. *Business Insights*.
3. *Business Optimization*.
4. *Data Monetization*.
5. *Business Metamorphosis*

b. Tata Kelola IT dengan *framework* COBIT 5

COBIT (*Control Objectives for Information and Related Technology*) adalah salah satu kerangka kerja (*framework*) yang diciptakan oleh ISACA yang berfungsi untuk dijadikan panduan dalam rangka mencapai tujuan organisasi. Organisasi yang dimaksud merupakan sebuah *enterprise* yakni organisasi yang menjalankan fungsi IT sebagai bagian dari proses bisnisnya.

Menurut COBIT, informasi merupakan sumber daya kunci bagi sebuah *enterprise*.

Teknologi mempunyai peranan penting terhadap proses kelangsungan informasi, mulai dari sebuah informasi diciptakan sampai dimusnahkan. *Enterprise* yang sukses memperlakukan TI sebagai bagian yang signifikan dalam melaksanakan proses bisnis. Proses bisnis dan TI harus berkolaborasi dan bekerjasama sehingga TI dapat masuk ke dalam tata kelola dan manajemen.

COBIT 5 memiliki beberapa pengungkit yang diturunkan dari tujuan organisasi yang telah didefinisikan. Pengungkit tersebut merupakan faktor yang mempengaruhi *governance* dan *management* dari *enterprise* TI, yaitu:

- 1) Prinsip, kebijakan dan *framework*;
- 2) Proses;
- 3) Struktur organisasi;
- 4) Budaya, etika dan perilaku;
- 5) Informasi;
- 6) Layanan, infrastruktur dan aplikasi; dan
- 7) SDM, kemampuan dan kompetensi.

Menurut COBIT 5, informasi yang efektif adalah informasi yang mampu memenuhi kebutuhan konsumen informasi (*stakeholder*). Dalam kasus *Big Data*, *enterprise* (organisasi) adalah *stakeholder* dan salah satu pilar utama adalah kualitas informasi. *Big Data harvesting* diharapkan bisa menghasilkan kualitas informasi yang mendukung pengambilan keputusan. Kualitas informasi yang baik akan mengakibatkan hasil keputusan organisasi yang baik sehingga akan meningkatkan keuntungan bagi *enterprise*.

Big Data merupakan sebuah proses pengumpulan data untuk menemukan pola dan korelasi yang mungkin tidak jelas pada awalnya, tetapi berpeluang menjadi berguna dalam pengambilan keputusan bisnis. Data tersebut seringkali merupakan data pribadi yang berguna yang dapat dikategorikan sebagai *Volunteered data*, *Observed data* ataupun *Inferred data* (Richard Chew, 2013).

Sumber Daya Manusia

Karena kompleksitas dari sistem *Big Data*, dibutuhkan kemampuan teknis beragam untuk dapat mengimplementasikannya. Namun karena *Big Data Engineering* sendiri masih merupakan sesuatu yang baru serta berhadapan dengan teknologi dan posisi pekerjaan yang juga baru, maka saat ini belum ada spesifikasi baku mengenai kompetensi SDM yang dibutuhkan untuk bidang ini. Berdasarkan proses kerja *Big Data* pada umumnya yaitu *Collect, Store, Transform, dan Analysis*³, maka ada empat hal yang harus diperhatikan terkait dengan kompetensi SDM yang dibutuhkan dalam implementasi *Big Data*, diantaranya:

1) Data Collection

Data yang akan diproses dalam sistem *Big Data* biasanya diambil dari website atau API (*Application program interface*), pada umumnya dengan menggunakan teknik *crawling*. Kompetensi SDM yang diperlukan antara lain:

- *Data APIs*
- *SQL dan Data Modeling*

2) Data Warehouse

Data yang telah diambil dari berbagai macam sumber akan disimpan dalam *server* yang sudah disiapkan untuk sistem *Big Data*. Sesuai dengan namanya, *Big Data* membutuhkan kapasitas penyimpanan data yang sangat besar, karena masuknya data yang sangat besar dan beragam ke dalam *server* setiap harinya. Salah satu alasan inilah disebut **Data Warehouse** karena proses penyimpanan, pengolahan, dan pengambilan data dari *server* akan sangat berbeda dengan **Database** biasa. Kompetensi yang dibutuhkan diantaranya ialah:

- *Relational Databases (e.g.: MySQL, MS SQL Server, Oracle, DB2, dsb)*
- *NoSQL (e.g.: HBase, SAP HANA, HDFS, Cassandra, MongoDB, CouchDB, Vertica, Greenplum, Pentaho, Teradata, dsb)*

3) Data Transformation

Salah satu sifat dari *Big Data* selain ukuran data yang sangat besar, juga jenis data yang sangat beragam. Agar data-data tersebut dapat dianalisis dengan baik, maka terkadang data-data tersebut perlu diubah ke dalam format lain sehingga memungkinkan untuk dianalisis. Kompetensi yang dibutuhkan termasuk:

- *ETL Tools (e.g.: Informatica, DataStage, SSIS, Redpoint, dsb)*
- *Scripting (e.g.: Linux/Unix commands, Python, Ruby, Perl, dsb)*

4) Data Analysis

Tahapan terakhir yaitu menganalisis data yang sudah diambil dan dikumpulkan sebelumnya, mengolah data tersebut menjadi informasi, hingga menjadi hasil statistik jika diperlukan. Kompetensi yang dibutuhkan diantaranya ialah:

- *MapReduce, Hadoop, Cloudera, IBM Big Insights, Hortonworks, MapR, dsb.*
- *Data mining atau machine learning (e.g.: Mahout, Neural Network, dsb)*
- *Statistical analysis software (e.g.: R, SPSS, SAS, Weka, MATLAB, dsb)*
- *Programming skills (e.g.: Java, Scala, Ruby, C++, dsb)*

Sistem

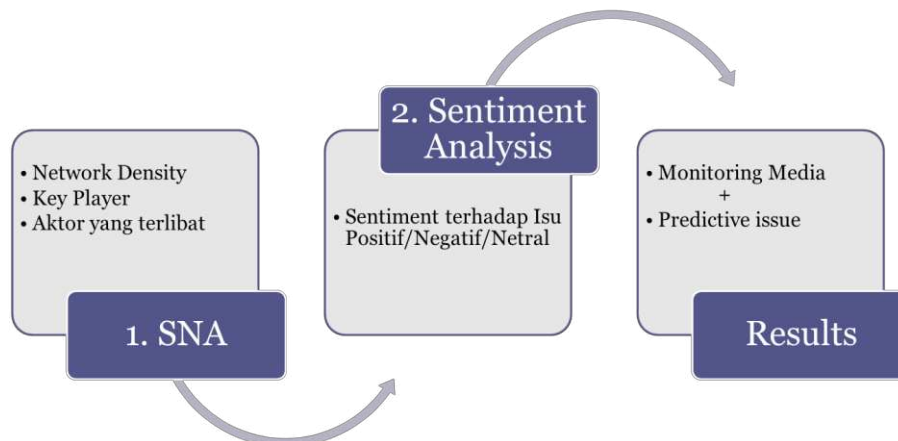
Jika *Big Data* diimplementasikan untuk monitoring media, maka yang akan menjadi inti dalam sistem tersebut adalah cara yang dipakai untuk menganalisis teksnya (*Text Analysis*). Ada banyak cara dan algoritma yang dapat dipergunakan untuk menganalisa teks, tergantung pada hasil seperti apa yang akan kita harapkan.

Salah satu proses dalam Monitoring Isu Publik adalah menganalisis isu dalam suatu berita. Salah satu metode yang bisa digunakan dalam menganalisis isu tersebut ialah dengan menggunakan kombinasi antara *Social Network Analysis* untuk melihat jaringan dari isu tersebut dan metode *Sentiment Analysis*, yang merupakan studi analisis opini, sentimen, evaluasi, pujian, sikap, serta emosi orang

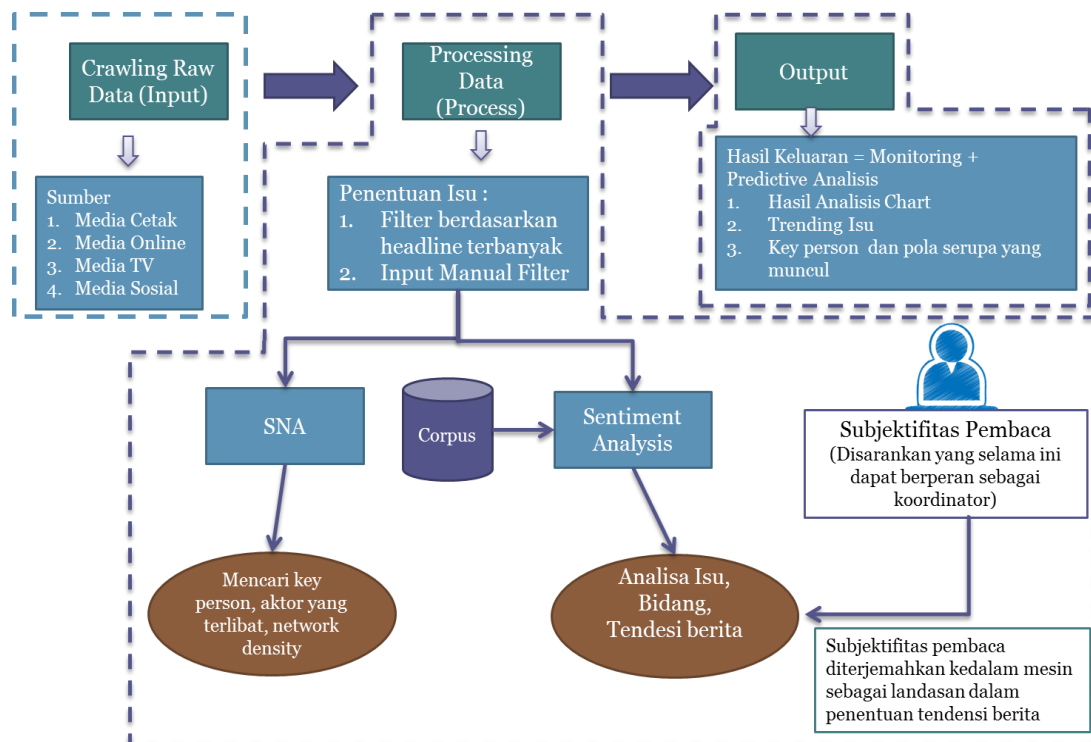
³ "The Key Skills Needed by Big Data Engineers". <http://insights.dice.com/2014/08/21/key-skills-needed-big-data-engineers/>. Diakses tanggal 19 Oktober 2016.

terhadap suatu entitas atau objek yang bisa berupa produk, jasa, individu, organisasi, kejadian, maupun sebuah topik. Penggunaan *Social Network Analysis* (SNA) lebih melihat kepada aktor yang terlibat dalam suatu isu. SNA bekerja dengan menganalisis *network* (jaringan) yang terbentuk antar aktor, sehingga

dapat terlihat aktor-aktor yang paling berpengaruh, hubungan antar aktor, pihak-pihak lain yang ada dalam jaringan tersebut, dan sebagainya. Adapun gambaran sederhana implementasi *Big Data* (kombinasi *sentiment analysis* dan SNA) dalam monitoring media ditunjukkan dalam Gambar 7 dan Gambar 8.



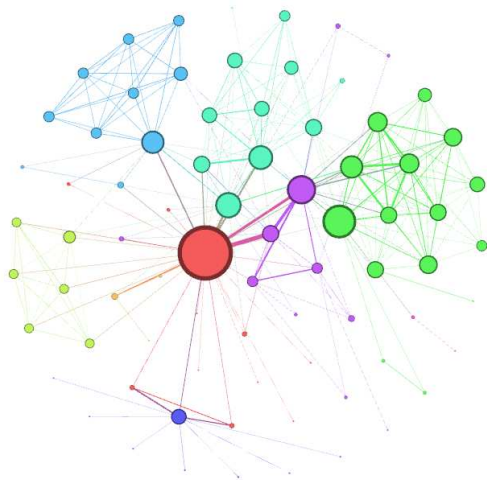
Gambar 7. Kombinasi *Sentiment Analysis* dan SNA



Gambar 8. Gambaran Sederhana kombinasi SNA dengan *Sentiment Analysis*

1. Social Network Analysis (SNA)

Penggunaan *Social Network Analysis* dapat diterapkan untuk melihat jaringan (*network*) yang tercipta antar aktor yang terlibat dalam suatu isu, seperti terlihat dalam ilustrasi Gambar 9.



Gambar 9. Ilustrasi SNA

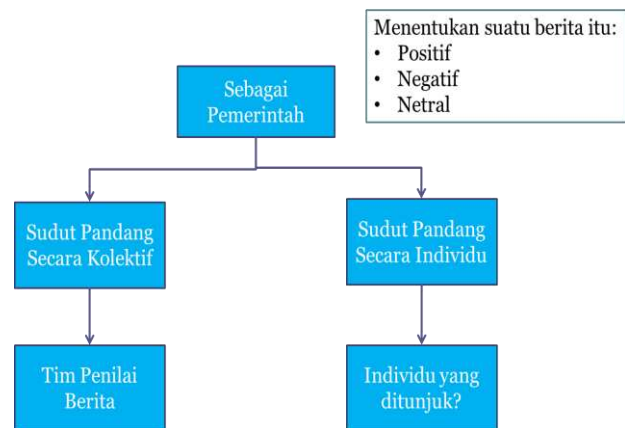
Social Network Analysis dapat mempresentasikan hal-hal kompleks, seperti berbagai peran aktor dalam jejaring sosial, konteks, komunitas, dan lain-lain. Penerapannya dalam Monitoring Media akan memungkinkan untuk melihat jaringan sosial yang terlibat dalam suatu isu tertentu. Melalui SNA, dua poin penting yang akan merubah skala dari suatu isu dapat terlihat, diantaranya :

- ✓ Jika isu yang sama dibicarakan oleh banyak orang dalam waktu yang singkat, maka isu tersebut berpotensi tinggi untuk menjadi isu yang berskala besar.
- ✓ Jika terdapat suatu isu yang dibicarakan oleh aktor yang memiliki jaringan dengan tingkat ketebalan (*density*) jaringan yang besar, maka kemungkinan besar isu tersebut akan dibicarakan oleh orang-orang yang ternasuk dalam jaringannya (misalnya *follower* pada *twitter*).

2. Sentiment Analysis

Salah satu hal terpenting sebelum melakukan *Sentiment Analysis* adalah dengan menetapkan subjektivitas pembaca berita, sebagai siapakah kita akan melihat suatu berita.

Dalam hal ini karena *user* terkait adalah Direktorat PPI pada Ditjen IKP, maka kita memposisikan diri sebagai pemerintah dalam melihat suatu berita. Suatu berita yang dipandang positif oleh pemerintah bisa jadi dipandang negatif oleh masyarakat. Begitupun sebaliknya. Karena itulah subjektivitas dalam melihat suatu berita perlu diperjelas sehingga dapat diterjemahkan ke dalam mesin.



Gambar 10. Subjektivitas Pembaca Berita

Terdapat dua hal penting yang harus diperhatikan, yaitu:

1. Sudut Pandang Kolektif

Sebagai pemerintah, sudut pandang yang dikumpulkan dari beberapa individu membentuk sudut pandang kolektif (misalnya dengan membentuk Tim Penilai Berita). Kesulitannya adalah sudut pandang yang harus disepakati bersama. Karena itu pemilihan anggota Tim yang kompeten dan representatif menjadi sangat penting, agar sudut pandang kolektif yang terbentuk benar-benar dapat mewakili sebagai pemerintah.

2. Sudut Pandang Secara Individu

Penggunaan sudut pandang individu akan jauh lebih mudah untuk diterjemahkan ke dalam bahasa mesin. Meskipun subjektivitas seseorang akan berbeda dengan yang lainnya, namun subjektivitas orang yang ditentukan tetap dapat diterima dan digunakan sebagai sudut

pandang pemerintah dalam melihat suatu berita.

Sentiment Analysis memungkinkan *user* untuk menganalisis suatu isu dengan memecah kalimat menjadi kata per kata. Metode *Sentiment Analysis* yang dapat digunakan dan sesuai dengan tugas dan fungsi Media Monitoring Kominfo diantaranya adalah:

- *Naïve Bayes Classifiers*

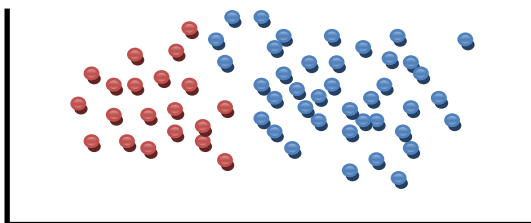
Naïve Bayes Classifiers merupakan metode yang berasal dari teori Bayesian. Model ini berdasarkan *conditional independency* yang dihasilkan dari prediksi kelas target. Berdasarkan teori Bayes, posterior probabilitas dapat dituliskan sebagai berikut:

$$p(C_k|\mathbf{x}) = \frac{p(C_k) p(\mathbf{x}|C_k)}{p(\mathbf{x})}$$

Secara bahasa, persamaan diatas dapat disederhanakan menjadi:

$$\text{posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}$$

Posterior Probability atau peluang yang muncul setelah dilakukan pengujian, dalam teori Bayes dihasilkan dari peluang awal dikalikan dengan kemungkinannya, dan dibagi dengan hasil pengujian. Ini merupakan metode pengelompokkan secara statistik yang digunakan untuk memprediksi peluang keanggotaan baru yang muncul dalam suatu kelompok. Dell juga mengeluarkan *textbook* tentang *Naïve Bayes Classifier* yang digambarkan secara sederhana.⁴



Gambar 11. Ilustrasi pengklasifikasian objek

Sebagai ilustrasi konsep *Naïve Bayes Classification*, objek yang ada dapat dilabelkan terlebih dahulu (Dell. 2015). Untuk memudahkan, ilustrasi objek diberi label *BLUE* dan *RED*. Dalam *Bayesian Analysis*, dikenal dengan *prior probability* (probabilitas awal). Karena jumlah *BLUE* dua kali lipat dari jumlah *RED*, maka dipercaya bahwa objek yang baru memiliki kemungkinan yang besar untuk memiliki label *BLUE*. Bisa juga dituliskan seperti:

$$\text{Prior probability for BLUE} = \frac{\text{Number of BLUE objects}}{\text{Total number of objects}}$$

$$\text{Prior probability for RED} = \frac{\text{Number of RED objects}}{\text{Total number of objects}}$$

Jika diterapkan dalam menganalisis media, maka tahap pertama yang dilakukan adalah dengan memberikan label pada teks-teks dalam *Corpus* yang digunakan (misal: negatif atau positif). Setelah pelabelan, dilakukan *Supervised Training* dengan menggunakan *dataset* yang dimiliki. Seperti halnya *Machine Learning*, semakin banyak dilatih maka akan semakin akurat hasilnya. Terutama pada *Naïve Bayes Classification*. Metode ini memiliki karakteristik yang jika digunakan pada data yang berskala besar akan mengeluarkan hasil yang lebih akurat.

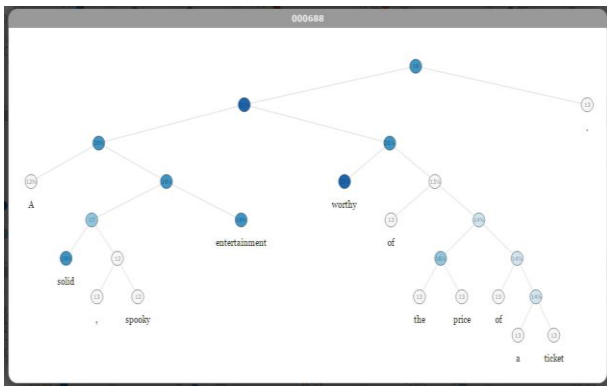
- *Standford's Sentiment Treebank*

Berbeda dengan *Sentiment Analysis* yang umumnya memecah kalimat dan melihat secara kata per kata, metode ini melihat teks dalam satu struktur kalimat yang utuh. Struktur kata dalam sebuah kalimat tidak dihilangkan, sehingga makna suatu kata dapat terlihat dalam satu kalimat yang utuh.

Kumpulan *words classifiers* dapat bekerja dengan baik pada kalimat yang panjang dengan mengandalkan kata-kata yang sederhana namun memiliki sentimen yang kuat seperti “hebat” atau “luar biasa”. Namun akurasi yang dihasilkan untuk klasifikasi satu

⁴ Statistics – Textbook, *Naïve Bayes Classifier*. Dell. [https://documents.software.dell.com/statistics/textbook/naive-bayes-classifier#technical notes](https://documents.software.dell.com/statistics/textbook/naive-bayes-classifier#technical%20notes)

kalimat dalam menentukan sentimen negatif atau positif tidak pernah melebihi 80%. Dalam kasus kalimat dengan *tone* netral untuk pesan pendek di *twitter*, akurasi yang didapat bahkan cenderung dibawah 60% (Wang et al., 2012)⁵



Gambar 12. Contoh hasil *Stanford's Sentiment Treebank*⁶

Pilihan Alternatif Implementasi *Big Data*

Metode apapun yang akan digunakan untuk analisis teks, *corpus* tetap menjadi unsur krusial. Hal ini terkait variasi *corpus* dalam Bahasa Indonesia yang masih sangat terbatas. Adapun *corpus* yang cukup lengkap biasanya dimiliki oleh *vendor* tertentu dan bersifat tertutup untuk umum. Pembuatan *Corpus* tersebut sangat penting, karena dapat menentukan keakuratan hasil *Sentiment Analysis*.

Tahap terakhir yang menjadi salah satu pertimbangan penting terkait bagaimana sistem *Big Data* tersebut akan diimplementasikan. Terdapat tiga opsi alternatif yang dapat dijadikan pertimbangan, antara lain:

1. Membangun Sistem Sendiri

Yaitu membangun sistem *Big Data* secara keseluruhan dari awal, meskipun dalam pengerjaannya dapat dilakukan dengan meng-hire pihak ketiga. Sebagai instansi

pemerintah yang memegang salah satu sektor strategis, yaitu informasi, pilihan untuk membangun sistem *Big Data* sendiri dapat menjadi pertimbangan. Keunggulan dari opsi ini antara lain :

- Membangun sistem dapat menjadi Investasi dalam pengelolaan informasi publik, yang akan bertambah kompleks seiring dengan berkembangnya media komunikasi
- Keamanan Informasi lebih terjamin karena sistem yang dibangun dimiliki sendiri
- Kedaulatan data di tangan pemerintah yang membangun sistem tersebut
- Biaya lebih terjangkau jika dibandingkan langganan *service* secara terus menerus
- Ketergantungan pemerintah terhadap pihak swasta berkurang, karena dikelola secara mandiri

Sedangkan kelemahannya antara lain :

- Pembangunannya memerlukan waktu panjang sampai bisa digunakan
- Memerlukan tenaga SDM dengan bidang keahlian tertentu
- Membutuhkan server yang besar untuk menyimpan dan mengolah data. *Big Data* memerlukan kemampuan dan kapasitas server yang dapat diandalkan agar data yang didapat bisa diolah dengan cepat dan akurat
- Membutuhkan kepastian dari pimpinan puncak untuk dapat menjamin diberikannya *support*, baik dari sisi tata kelola, SDM, maupun biaya
- Harus ada seseorang yang mengawal proses pembuatan dari awal hingga selesai agar hasil akhir dapat sesuai
- Perlu penyediaan infrastruktur yang memadai (Jaringan, *Server*, dsb)

2. Berlangganan

Saat ini sudah ada beberapa perusahaan/*vendor* yang menawarkan *service Big Data*. *Vendor* yang *notable* diantaranya:

- *i811*
- *Indonesia Indicator/ebDesk*
- *Mediatrac*

⁵ S. Richard, P. Alex, W. Y. Jean, C. Jason. *Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank*

⁶ *Screenshot* diambil dari hasil uji coba di website *Treebank* Stanford. <http://nlp.stanford.edu/sentiment/treebank.html>. Diakses tanggal 15 Juni 2016

- *Maverick*
- *Mediawave*
- *Insentia*
- *Awesometrics*
- *IMMC (Indonesia Media Monitoring Center)*
- *IMM (Intelligence Media Management)*

Meskipun terdapat kekurangan pada masing-masing vendor, pilihan untuk berlangganan tetap dapat menjadi opsi jika layanan yang diberikan dapat di *custom* sesuai dengan kebutuhan. Keunggulan opsi ini antara lain :

- Pengelolaan dan reliabilitas sistem dijamin oleh penyedia layanan
- SDM yang diperlukan hanya sebagai pengguna, sehingga tidak diperlukan kualifikasi SDM teknis yang tinggi
- Sistem dapat langsung digunakan
- Tidak membutuhkan *server* yang besar karena pemrosesan data di pihak penyedia layanan

Sedangkan kelemahannya antara lain:

- Biaya lebih tinggi (Ditentukan oleh pihak penyedia, estimasi untuk layanan yang pernah digunakan seperti IMM sekitar 30 Juta per bulan)
- Keberlangsungan sistem ditentukan oleh penyedia layanan, sehingga menyebabkan ketergantungan terhadap vendor
- *Raw data*, *algoritma*, dan *corpus*, dipegang oleh penyedia layanan, karena hal tersebut merupakan *core* yang diperdagangkan.
- Membutuhkan penyesuaian sistem yang terperinci agar layanan sistem Big Data yang diberikan sesuai dengan kebutuhan. Karena akan berbeda penggunaan dan kebutuhan antara *corporate* dengan pemerintah.

3. Membangun Sebagian

Pemerintah membangun bagian dari sistem yang memiliki nilai strategis bagi pemerintah (misalnya seperti *Corpus*), sedangkan sisanya dibangun pihak luar. Keunggulan opsi ini antara lain :

- Dari sisi keamanan informasi, bagian sistem yang penting dapat dimiliki sendiri (e.g.: Kedaulatan Data, *Corpus*, dsb)
- Sistem dapat dibuat sesuai kebutuhan/ keinginan
- Sebagai pengguna akan lebih memahami sistem, karena mengikuti proses pembuatan dari awal hingga jadi
- Tidak memerlukan banyak SDM teknis, karena pengerjaannya akan di-*share* dengan pihak luar

Sedangkan kelemahannya antara lain:

- Harus ada seseorang yang mengawal proses pembuatan dari awal hingga selesai agar hasil akhir dapat sesuai dengan harapan
- Membutuhkan kepastian dari pimpinan puncak untuk dapat menjamin diberikannya *support*, baik dari sisi tata kelola, SDM, maupun biaya
- Perlu penyediaan infrastruktur yang memadai (*Jaringan*, *Server*, dsb)

Karena pengerjaannya merupakan kerjasama antara pemerintah dengan pihak luar, maka perlu adanya suatu komitmen tertulis yang dibuat terperinci untuk memastikan sistem yang dibuat bersama dapat selesai dan terhindar dari masalah-masalah yang dapat muncul kedepannya.

PENUTUP

Simpulan

Dalam sistem *Big Data* yang akan diimplementasikan untuk media monitoring, *Text Mining* dengan metode *Sentiment Analysis* akan menjadi *core* dalam sistem tersebut. Terdapat dua metode *Sentiment Analysis* yang paling sesuai untuk diterapkan, yaitu *Naïve Bayes Classifiers* dan *Stanford Treebank*.

Metode *Sentiment Analysis* yang paling sesuai bagi media monitoring adalah *Stanford Treebank*, karena metode tersebut dapat melihat hubungan antar kata dalam tiap kalimat, berbeda dengan metode lain yang umumnya memaknai kalimat dengan memecah

kata-perkata. Namun karena metode tersebut masih dalam tahap pembuatan dan *Corpus* yang digunakan dalam bahasa Inggris, maka *Naïve Bayes* dapat menjadi pilihan untuk diterapkan bagi monitoring media saat ini.

Terdapat tiga alternatif terkait tahapan implementasi sistem *Big Data* yang dapat dilakukan, antara lain: membangun sistem sendiri, berlangganan yang sudah ada, atau membangun sistem sebagian. Masing-masing alternatif tersebut memiliki keunggulan dan kelemahan yang perlu diperhatikan untuk menjadi pertimbangan.

Saran

Metode *Sentiment Analysis* sebaiknya dikombinasikan dengan *Social Network Analysis*, agar *output* yang dihasilkan tidak hanya berhenti pada monitoring media, tetapi hingga menemukan *key person* atau aktor yang terkait pada isu-isu tersebut. Akan sangat disayangkan jika *resource* besar yang telah dialokasikan hanya dipergunakan untuk monitoring media. Pemanfaatan *Big Data* dapat lebih optimal jika diperluas untuk melakukan *predictive analysis* dalam memperkirakan potensi suatu isu menjadi besar. Kedepannya melalui sistem *Big Data*, Kominfo bersama instansi pemerintah lain dapat mengantisipasi dan bereaksi lebih cepat terhadap isu yang diprediksikan menjadi besar, sehingga dapat dikendalikan.

Proses dalam sistem *Big Data* yang akan diimplementasikan secara sederhana dapat terbagi dalam tiga kategori, yaitu ***Input, Process, dan Output***. Kombinasi metode yang direkomendasikan ada pada tahap *process*. Diharapkan rekomendasi ini dapat memaksimalkan fungsi *Big Data* tanpa mengubah fungsi utama dari monitoring media. Implementasi *Big Data* untuk mendukung monitoring media sebaiknya dilakukan dengan membangun sistem sebagian, sehingga bagian sistem yang penting seperti *Corpus* dapat dimiliki sendiri, serta sistem yang dibangun dapat dirancang dan diawasi sesuai kebutuhan. Opsi ini dapat mengurangi ketergantungan pemerintah terhadap pihak lain, juga menekan

alokasi dana yang besar untuk vendor tiap tahunnya.

Pembuatan *Corpus* sendiri dapat dilakukan secara independen ataupun bekerja sama dengan pihak Universitas, sehingga pemerintah dan akademisi dapat saling berbagi dalam hal pengetahuan dan sumber daya (*resources*).

Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Puslitbang Aptika dan IKP, Badan Litbang SDM, Kominfo, dimana tulisan ini merupakan bagian dari penelitian Tim yang pernah dilakukan di Puslitbang Aptika dan IKP dengan judul “**Studi Implementasi Sistem Big Data untuk Mendukung Kebijakan Komunikasi dan Informatika**” yang dibiayai dari DIPA tahun 2016 Puslitbang Aptika dan IKP. Penulis juga mengucapkan terima kasih kepada tim penelitian ini yang terdiri dari peneliti di Puslitbang Aptika dan IKP dan tim dari Telkom University.

DAFTAR PUSTAKA

- Ashari, Mochamad, 2015. *Big Data Industry and Academic Point of View*. Konferensi Big Data Indonesia.
- IBM Big Data & Analytics Hub, 2013. *4v's of Big Data*. www.ibmbigdatahub.com/infographic/four-vs-big-data. Diakses tanggal 17 Februari 2016.
- Instruksi Presiden No. 9 Tahun 2015 tentang Pengelolaan Komunikasi Publik.
- Krishnan, K., 2013. *Data Warehousing in the Age of Big Data*. USA: MK Publications.
- Liu, Z., Ping, Y., Lixiao, Z., 2013. *A Sketch of Big Data Technologies*. Seventh International Conference on Internet Computing for Engineering and Science. School of Information Science and Technology, Shanghai Sanda University Shanghai, China.

- Matsudaira, Kate, 2014. “*The Key Skills Needed by Big Data Engineers*”.
<http://insights.dice.com/2014/08/21/key-skills-needed-big-data-engineers/>. Diakses tanggal 19 Oktober 2016.
- Morrison, C. W., Andi, & H., Farid, 2010. *Teori Komunikasi Massa-Media, Budaya dan Masyarakat*. Jakarta: Ghalia Indonesia.
- Munassar, Nabil, M.A., & A., Govhardan, 2010. *A comparison Between Five Models Of Software engineering*. IJCSI International Journal of Computer Science Issue, Vol.7, pp.95.
- Paparan PT. Delapan Sebelas Indonesia tentang Pamedi. 2 September 2015
- Prajapati, V., 2013. *Big Data Analytics with R and Hadoop*. Birmingham, UK: Packt Publishing.
- Romney, Marshall, B. & Paul, J.S, 2005. *Accounting Information Systems (9th Edition)*. Jakarta: Salemba Empat.
- Sagiroglu, Seref, & Duygu, S., 2013. *Big Data: A Review*. IEEE International Congress on Big Data, Gazi University, Department of Computer Engineering, Faculty of Engineering, pp. 42-47.
- Sawant, N., & Himanshu, S., 2013. *Big Data Application Architecture Q & A*. New York: Springer Science Business Media.
- Schell, R., 2013. *Security – A Big Question for Big Data*. IEEE International Conference on Big Data University of Southern California, USA.
- Schmarzo, Bill, 2013. *Understanding How Data Powers Big Business*. USA: John Wiley & Sons, Inc.
- Schonberger, V.M., & Kenneth, N.C., 2013. *Big Data: A Revolution that Will Transform How we Live, Work, And Think*. New York, USA: Houghton Mifflin, Harcourt Publishing.
- Sommerville, Ian, 2011. *Software engineering 9th Edition*. Boston: Addison-Wesley Publishing Company.
- Statistics – Textbook, *Naïve Bayes Classifier*. Dell.
[https://documents.software.dell.com/statistics/textbook/naive-bayes-classifier#technical notes](https://documents.software.dell.com/statistics/textbook/naive-bayes-classifier#technical%20notes). Diakses tanggal 15 Juni 2016.
- Subiakto, Henry & Rachmad, I., 2014. *Komunikasi Politik, Media dan Demokrasi (Cetakan Kedua)*. Jakarta: Kencana Prenadamedia Group.
- T., Firat, & Keane, J.A., 2013. *Big Data Framework*. IEEE International Conference on Systems, Man, and Cybernetics. School of Computer Science, The University of Manchester, UK. pp. 1494-1499.
- Undang-Undang Nomor 14 Tahun 2008 tentang Keterbukaan Informasi Publik.
- Zikopoulos, Paul, C., 2013. *The Power of Big Data : The IBM Big Data Platform*. USA: Mc Graw Hill.

